

Appendices for:

Levels of Analysis for the Study of  
Environmental Health Disparities

M. Soobader<sup>1</sup>  
C. Cubbin<sup>2</sup>  
G.C. Gee<sup>3</sup>  
A. Rosenbaum<sup>4</sup>  
J. Laurenson<sup>4</sup>

---

<sup>1</sup> Address correspondence to: Mah-J Soobader, STATWORKS, 800 Matthew Court, Suite 102, Braintree, MA 02184, USA. Tel: (781)843-0821. Email: [Mah-J@statworks.com](mailto:Mah-J@statworks.com)

<sup>2</sup> Center on Social Disparities in Health, Department of Family and Community Medicine, University of California, San Francisco, CA, USA

<sup>3</sup> University of Michigan School of Public Health, Department of Health Behavior and Health Education, Ann Arbor, MI, USA

<sup>4</sup> ICF Consulting, Fairfax, VA, USA

## **Appendix A. Multilevel Modeling Approaches for Environmental Health Disparities**

Appendix A has been adapted from the writings of Subramanian (2003), Duncan (1996; 1998), and Jones (1991), who have provided leadership in applying multilevel models to the field of social epidemiology. This Appendix was written for conceptual understanding and, therefore, mathematical and statistical details are minimized (for more technical discussions, see Bingenheimer and Raudenbush, 2004; Bryk and Raudenbush, 1992; De Leeuw and Kreft, 2001; Goldstein, 1995; Hox and Kreft, 1994; Singer, 1998). We use the word “context” as a generic description of places or areas that may be administratively defined (as in census tracts, counties, and states) or socially defined, as in neighborhoods and communities. The notion of context can also be extended to temporal contexts, as in different time periods. Throughout Appendix A we use a worked example of individual lead exposure to provide a conceptual modeling approach for:

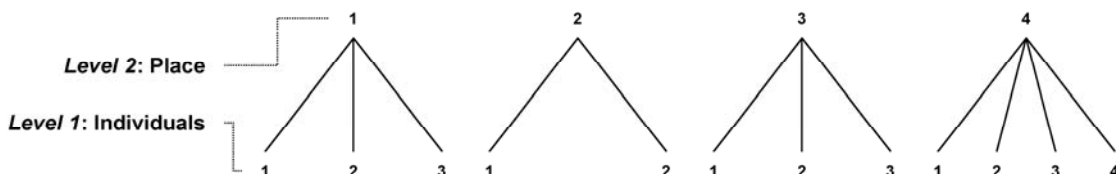
- Examining a single environmental exposure that may occur through multiple media operating at different levels simultaneously and interacting at different levels.
- Examining multiple exposures operating at different levels simultaneously, potentially accumulating over time, and interacting with each other.
- Examining exposures differentially affecting subgroups of the population and/or geographic areas, and/or producing synergistic outcomes.
- Examining the fundamental role of social and economic factors and the need to account for all levels through which these mechanisms influence individual exposures, either directly or through their effect on local environments.

In the following sections, we discuss these issues in greater detail, illustrating each with a sample research question in a text box.

## Composition and Context (Partitioning the Variance)

*Research Question: Are there contextual differences in individual lead levels between contexts, after taking into account individual characteristics?*

Multilevel models provide the advantage of identifying and differentiating sources of variation at multiple levels, thus assigning variability to the appropriate level. It is well established that environmental exposures vary by place. What is unclear is whether this variation is attributable to the composition of individuals living in that area or if these variations are independent place variations. Therefore, in starting to examine individuals within contexts, we need to identify which effects are compositional (i.e., due to the characteristics of individuals) and which are contextual (i.e., due to the characteristics of places). In its simplest form (see Figure A1), a two-level multilevel model of individuals within places would allow us to differentiate between contextual (geographic/place) sources of variation, compositional (individual) sources of variation, and the variation due to the interaction between composition and context.



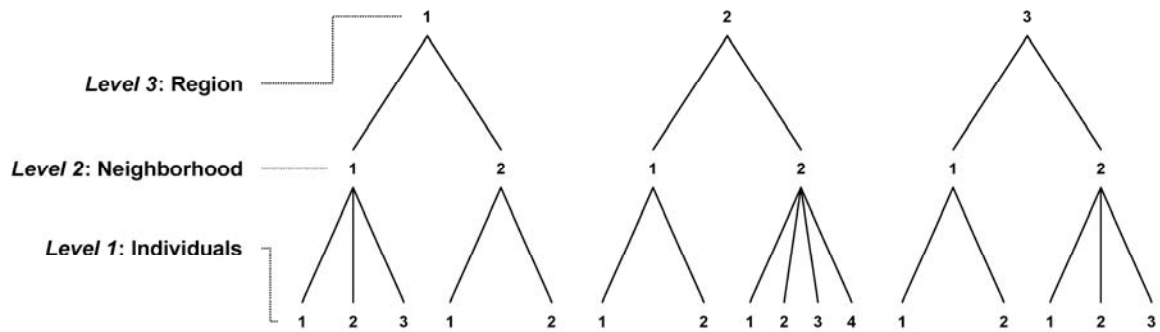
**Figure A1: Two-level structure. (Source: Duncan et al., 1996)**

Environmental studies are typically conducted at a single level, either at the aggregate/ecologic level or the micro/individual level. These studies have been criticized due to the incorrect inference of the study results. The former relates to the well-known aggregate/ecologic fallacy when analyses are only done at the aggregate level and inference is made to the individual; the individualistic/atomistic fallacy occurs when analyses are done only at the individual level and inference is made to the group (Diez-Roux, 1998). But multilevel level models are not just about avoiding fallacies; they also can provide insight into the complex processes that influence health.

### **Multiple Contexts at Different Levels**

*Research Question: What levels are important for the study of lead exposure, and what is the relative importance of the different levels?*

Since multilevel models are not restricted to just two levels and potentially can be expanded to  $n$  levels, these models allow us to explore the importance of all relevant levels. For example, when the same exposure is measured at multiple levels these models allow us to evaluate the relative importance of the exposure at each level. Multilevel models summarize the variability between higher-level units, such as the variability between neighborhoods within counties (Rice and Jones, 1997). One may explore the variability of lead exposure at the neighborhood (e.g., census tract), county, and state level to ascertain which level is most important in contributing to the observed variation in individual lead levels (see Figure A2).



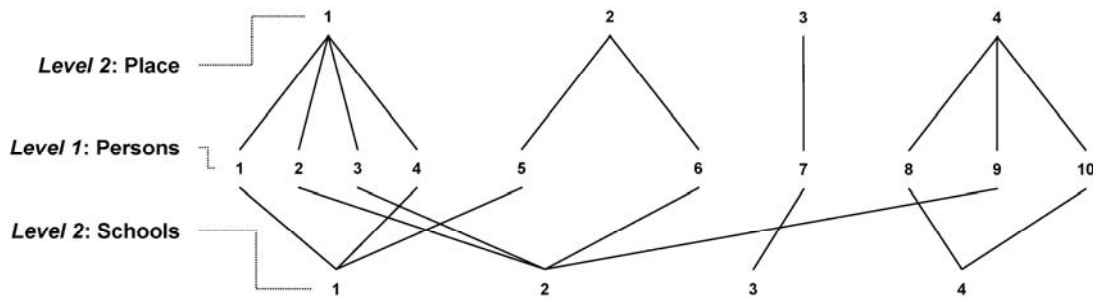
**Figure A2: Three-level structure.** (Source: Duncan et al., 1996)

Such an approach may also be important when examining the role of environmental policy and its implementation at the state, county, and local level, since states vary in their regulation and resource allocation for hazards such as lead exposure (Brown et al., 2001).

### Multiple Contexts at the Same Level

*Research Question: What is the relevant contribution of neighborhood and school levels, that may not be nested within one another but overlap, to lead exposure?*

An important feature of multilevel models is that the data need not be hierarchical. That is, contexts do not need to be neatly nested within each other. This is important as exposures commonly occur in contexts that are not hierarchical, but different contexts may occur at the same level. For example, children may be exposed to lead in the neighborhood but also within the school environment, and neighborhoods may not be nested within school districts. In this situation we find that a number of different contexts may overlap at the same level. Such contexts are referred to as cross-classified structures. Figure A3 illustrates this concept, with individuals at level 1 nested within both neighborhoods and schools at level 2.



**Figure A3: Multiple contexts at the same level. (Source: Duncan et al., 1996)**

This is important to identify contexts that may be having a confounding effect. For example, we may find that the observed variation between neighborhoods is actually variation between schools (Duncan et al., 1996).

## Interaction Effects

An important application of multilevel models is that it allows us to examine how variables measured at one level affect associations found at another (Bryk and Raudenbush, 1992). Multilevel models allow us to examine individual and contextual interactions as well as interactions between different levels of context. These interactions are of great concern for cumulative risk studies, as in determining whether the effects of two or more exposures are merely additive or synergistic. For example, will (local-level) air pollution potentiate the effects of (individual-level) lead exposure on childhood learning disabilities? That is, among individuals with the same lead levels, is the dose response of lead amplified among those living in high-smog neighborhoods compared with those living in low-smog neighborhoods?

### **Individual Contextual Interaction**

*Research Question: What is the average association between individual lead exposure and neighborhood quality, and does this association differ for different individuals based on their poverty profile, after accounting for individual characteristics and the neighborhoods in which the individuals live? That is, are poor persons living in low quality neighborhoods at higher risk of lead exposure than poor persons living in high quality neighborhoods?*

Including an interaction between the individual and the context provides information about the differential effect of context across individual groups; i.e., the characteristics of individuals and of places interact to produce different effects on individual blood lead levels. Extending our current example, we may now introduce neighborhood quality and examine its association with individual poverty. We may observe that poor individuals may experience different levels of lead exposure depending on the quality of the neighborhood in which they live.

### **Contextual and Contextual Interaction**

*Research Question: What is the average association between resource allocation at the city level and neighborhood quality in relation to lead exposure, and does this association differ for neighborhoods based on their quality profile, after accounting for individual characteristics and the characteristics of the neighborhoods in which the individuals live?*

Interactions may also be examined between different levels of context. Extending our example to include the level of cities with neighborhoods nested within cities, we can examine the effect of a city level variable on different types of neighborhoods. For example, including a

measure of resource allocation for preventing lead exposure at the city level, we may find that for a given amount of spending, we get better results for high quality neighborhoods compared to low quality neighborhoods. Such observations are important for policy development and resource allocation for preventing environmental exposures.

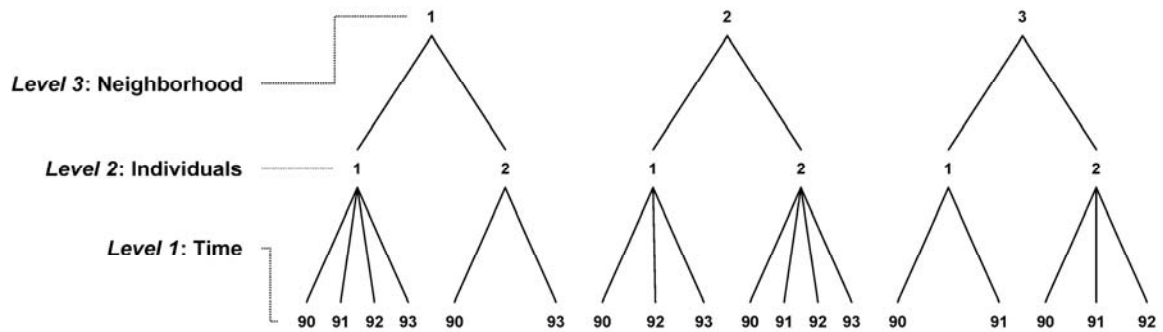
### **Modeling Contexts and Individuals Over Time**

Multilevel models allow us to examine changes over time, an important aspect in monitoring environmental health disparities. As contexts change over time, so do the exposure levels in individuals. There are two possible situations depending on whether individuals are repeatedly measured or the context is repeatedly measured. While the use of multilevel analysis of individuals nested within contexts is fairly intuitive, the repeated measurement of contexts is less so.

#### ***Repeated Measures of Individuals***

*Research Question: While individual lead exposures may have declined over time, have neighborhood contextual disparities declined or increased, and for which population groups have the contextual disparities declined or increased?*

If individuals are measured repeatedly over time, as in a panel design, their measurements can be described as being nested within each individual. For example, we may monitor lead levels of a group of individuals over time, so we would have each lead level for each time period at level 1 nested within individuals at level 2 nested within neighborhoods at level 3, as seen in Figure A4.



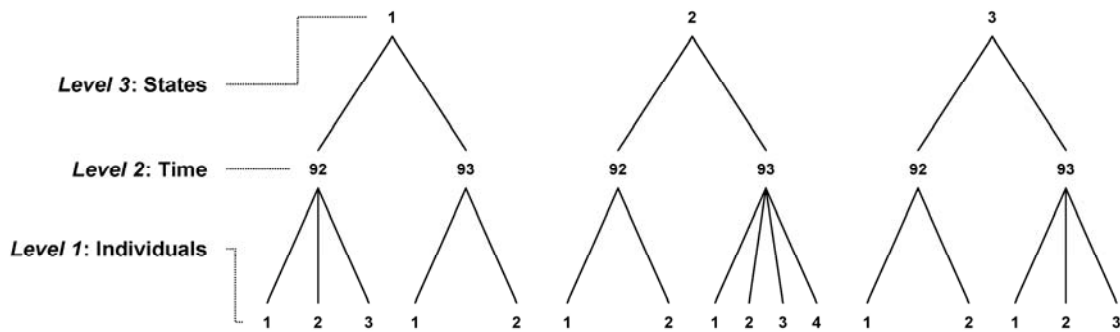
**Figure A4: Repeated measures of individuals.** (Source: Duncan et al., 1996)

The advantage of the multilevel design over conventional repeated measures analyses is that the number of measurements per individual, as well as the spacing between measurements, may vary. In this case we are examining individual change within a contextual setting (Duncan et al., 1996).

### ***Repeated Measures of Contexts***

*Research Question: Which types of individuals and which types of places have changed over time with respect to lead exposure?*

If repeated cross-sectional surveys are conducted within a certain context, they can be regarded as repeatedly measuring contexts over time. This design can be described as individuals nested within time, nested within contexts. For example, statewide surveys that are conducted annually will produce individuals nested within time nested within states as shown in Figure A5. In this case, we can examine trends within states while controlling for individual characteristics.



**Figure A5: Repeated measures of contexts. (Source: Duncan et al., 1996)**

Finally, multilevel models require that the spatial autocorrelation (the similarity between individuals for a given variable as a function of spatial distance) be accounted for, e.g., as in point sources of pollution. The combination of spatial models with multilevel models is relatively new, but published studies that incorporate structures that describe spatial adjacency are already available (Burnett et al., 2001; Langford et al., 1998).

### **Types of Contextual Variables Used in Multilevel Models**

While we have discussed the use of contextual variables, it is important to note that contextual variables can be measured and interpreted in different ways; this section introduces some commonly defined contextual variables. These variables are typically used to describe the characteristics of a group or context.

Derived variables are contextual variables that are summarized from the characteristics of individuals in that context, such as median neighborhood income or the percentage of high school graduates in a neighborhood. Some derived variables have no individual level equivalent, such as inequalities in the income distribution in an area, while others, such as average neighborhood income, do. While derived variables may be summarized from individual characteristics, their effect may be independent in that, conceptually, they may be measuring a

characteristic of the context. The term “derived variable” is used synonymously with analytic and aggregate variables (Diez-Roux, 1998; Diez Roux, 2002).

Integral variables, in contrast to derived variables, do not have an individual equivalent. Zoning policies, racial segregation, and population density are examples of integral variables. Thus, integral variables often describe group properties that are distinct from properties of the individuals comprising these groups (Diez-Roux, 1998; Diez Roux, 2002).

Environmental variables within the context of multilevel models have been described as measures of physical and chemical exposures. Environmental variables are generally not aggregated from individual-level variables but do have individual-level equivalents. Such variables are typically used as proxies for individual-level variables that may be difficult to measure at the individual level (Diez-Roux, 1998; Diez Roux, 2002). A common example of an environmental variable used as a proxy for an individual variable is the use of the ambient outdoor concentration of an air pollutant as a proxy for the personal exposure concentration of the local residents. Because (a) most people spend more than 80% of their time indoors or in vehicles, where pollutant concentrations can be significantly different from those outdoors, and (b) many people spend a substantial amount of time at locations other than their residence, the personal exposure concentration can be quite different from the ambient outdoor concentration in the vicinity of the residence. As explained in Appendix D, modeling tools and databases are available to estimate personal exposure concentrations. Multilevel methods advocate measuring the exposure at both the level of the residence and the level of the individual.

This Appendix has provided a brief overview of multilevel models. While multilevel models can provide an important tool to improve our understanding of environmental exposure

and its relationship to socioeconomic position, the primary research needs to be theoretically justifiable, and model complexity needs to be balanced with functional applications.

## **Appendix B: Place and Social Theory**

Many theoretical frameworks exist for examining the roles of social and physical environments in racial/ethnic and socioeconomic health disparities. These frameworks vary in name, disciplinary origin, and emphasis, but they share some common themes.

Several theoretical frameworks for understanding environmental health disparities have recently been proposed. Gee and Payne-Sturges (2004) build upon the exposure-disease paradigm in suggesting that psychosocial stress is the key component that explains the greater susceptibility of disadvantaged populations to environmental hazards. Calling it the stress-exposure disease framework, they hypothesize that residential segregation is the reason that “race” is important, incorporate a multilevel perspective, and argue that racial differences in stressors account for racial differences in vulnerability (Gee and Payne-Sturges, 2004). Schulz and Northridge (2004) have developed a framework for understanding social and environmental inequalities in health, drawing on earlier frameworks intended to understand racial disparities in health and incorporating factors in the built environment. In this framework, macro factors (e.g., structural determinants such as distribution of wealth) influence and are influenced by local factors (e.g., land use and community investment), which then influence and are influenced by proximate factors (e.g., health behaviors and housing conditions), which ultimately influence health and well-being (Schulz and Northridge, 2004). Morello-Frosch and colleagues (2002) suggest that income inequality and social capital at the macro levels affect the ability of local communities to influence environmental and social policies and, consequently, their ability to resist environmental health stressors such as the placement of hazardous waste facilities and subsequent health effects.

SEP is a key variable in the theories and frameworks summarized above. For example, a recent paper by O'Neill and colleagues (2003) illustrates how social epidemiology theory can be incorporated in understanding disparities in air pollution: "First, groups with lower SEP may have higher exposure to air pollution. Second, because lower-SEP groups already experience compromised health status... they may be more susceptible to the health effects of air pollution. Third, because of the combination of greater exposure and susceptibility, these groups are likely to suffer greater health effects from air pollution exposure." Racial/ethnic environmental health disparities can be viewed in a similar manner.

In modern social epidemiology, prominent theoretical models include psychosocial theory, social production of health, and ecosocial theory and its related frameworks (Krieger et al., 2001). Cassel, a leader in defining psychosocial theory, proposed that psychosocial factors in the social environment (e.g., isolation, disorganization, and support) influence the degree to which some persons become more or less susceptible to disease-causing agents than others (Cassel, 1976). In this model, the social environment affects health indirectly by influencing susceptibility through changes in the neuroendocrine function. Others have proposed that psychosocial factors can affect health directly through allostatic load, the "wear and tear" of organ systems resulting from stressors (McEwen, 1998). The social production of health perspective, sharing much in common with the political economy of health theory, stresses "upstream" political and economic determinants of health, including income inequality, racial discrimination, neoliberal economic policies, and deregulation of corporations (Krieger et al., 2001). In part a reaction to proponents of individual responsibility for health, the social production of health theory claims that class inequalities are the fundamental causes of health inequalities. Ecosocial theory (Krieger, 1994; Krieger et al., 2001) and its related multilevel

frameworks incorporate multiple levels of organization (biological, social, and ecological) over time and space to explain changing population patterns of health (Krieger et al., 2001). The intent of ecosocial theory is to provide a set of guiding principles for scientific inquiry and to incorporate accountability (e.g., institutions), as does the social production of health theory.

There are several common themes in the theories and frameworks discussed above. Of note are: (1) the concept of vulnerability, (2) multiple levels/nested hierarchies, and (3) the incorporation of time. “Vulnerability” is a defining concept in the field of environmental justice and has been categorized into at least four overlapping types: susceptibility/sensitivity (e.g., vulnerable populations such as children or the elderly), differential exposure (e.g., proximity to pollution sources), differential preparedness (e.g., low income), and differential ability to recover (e.g., discrimination) (Subramanian, 2004). Vulnerability is key to understanding both racial/ethnic and socioeconomic disparities in environmental health.

In addition, these theories and frameworks consider multiple levels and/or nested hierarchies. In psychosocial theory, emphasis is placed on at least two levels, the characteristics of individuals and of the social environment. The social production of health theory emphasizes the role of economic and political structural determinants of individual-level health, focusing on power relations and accountability. Ecosocial theory very explicitly includes consideration of multiple levels corresponding to proposed causal pathways (e.g., individual and neighborhood) and nested hierarchies (e.g., individuals within neighborhoods within cities within states). By definition, theoretical frameworks for investigating environmental health disparities include multiple levels and characteristics of environments and of those residing within them.

Time is an important concept for understanding environmental health disparities. Certain toxicants may have a greater adverse effect during certain ages, as with childhood lead poisoning

or fetal exposure to alcohol. These developmental effects have been characterized as “windows of vulnerability” (Dietert et al., 2000; Luster et al., 2003). Further, hazards may flux with time, such as seasonality of weather and temperature (Bhattacharya et al., 2003). From a lifecourse perspective, the accumulation of exposures to socioeconomic disadvantage/advantage or socioeconomic characteristics measured at critical time periods (e.g., childhood) could have an important influence on later health outcomes (Galobardes et al., 2004). In addition, relationships between various factors at multiple levels and health are not static. Characteristics of people and places change over time and should be modeled accordingly.

Recent work in social epidemiology discusses the multiple pathways through which socioeconomic characteristics of local places (i.e., neighborhoods) could potentially affect health (Macintyre et al., 2002). One is through the physical environment, including air/water/housing quality, affordable and nutritious food, and safe places to play/exercise. Another pathway is through the social environment, including processes such as social cohesion or the level of mutual trust among neighbors (Sampson et al., 1997); crime; acceptability of behaviors such as smoking, teen parenting, and adult monitoring of youth; and neighborhood reputation (Macintyre et al., 2002). A third pathway is through the service environment, including fire/police protection, access to health services, transportation, and other social services (e.g., education and job training/placement). Differences in local-level physical, social, and service environments could influence an individual’s health through behaviors such as smoking and health care use; psychological stressors such as fear or feeling deprived; hazardous exposures such as pollution, violence, or traffic; and/or opportunities for socioeconomic attainment such as availability of good schools and jobs. Macro-level factors, in turn, directly influence more local-level factors.

These multiple pathways have relevance for studying and monitoring environmental health disparities, since they directly impact health or influence vulnerability to environmental hazards.

Similar to studying and monitoring health disparities in general, research on environmental health disparities is rooted in the ethical principle of social justice or equity. Of primary interest is whether populations that are *a priori* socially disadvantaged in society (by nature of their SEP, race or ethnic group, gender, religious affiliation, etc.) are further—and unacceptably—disadvantaged with regards to their health (Braveman, 2003). Anchoring the study of environmental health disparities in a social justice framework has important operational implications. For example, in investigating disparities, a social justice framework would argue that the reference population (to which other groups are compared) would be the *a priori* most advantaged group (e.g., whites and the highest SEP group) rather than the population average or the group with the lowest risk (Braveman, 2003). Finally, a social/environmental justice approach implies that interventions should be aimed towards health promotion and sustainability rather than remediation only (Schulz and Northridge, 2004).

## **Appendix C: Census Data and Census Geography**

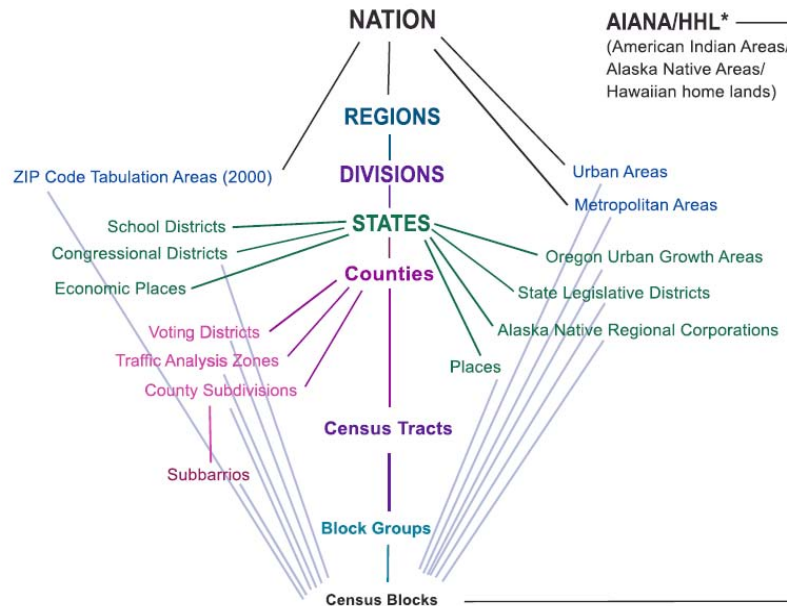
### **Census Data**

Census data are the most commonly used source for characterizing and defining contexts. The US Census is generally perceived to be of high quality, given its periodic nature, methodological sophistication, and the relatively small sampling errors. However, one area of concern, especially with regard to racial health disparities, is the well-documented undercount, especially of minority groups. In 2000, estimates of the nationwide undercount range from 0.12 to 1.15%, while the undercount for black males ranged from 2.1 to 7.7%. The magnitude of the undercount is a less serious consideration than the fact that the undercount varies differentially by race/ethnicity, and home ownership. Various methods for accounting for the undercount have been developed (Robinson, 2001; Williams et al., 2001). To overcome the limitation of census data being provided decennially, The American Community Survey will provide comparable data to that available on the US Census on an annual basis for all states, cities, counties, and metropolitan areas. For smaller areas such as census tracts, this survey will release estimates every 5 years (US Census Bureau).

### **Most Commonly Used Levels**

Census blocks are the smallest unit for which the census collects and tabulates data representing approximately 85 individuals (see Figure C1). Visible physical (streets, railroads, and streams) and cultural features (e.g., schools and other buildings) define census blocks. In 1990 the census provided tabulated block data for the entire US, recognizing the utility of these data for small area studies (US Department of Commerce et al., 1994).

**Figure C1: Standard Hierarchy of Census Geographic Entities**



Source: (US Census Bureau)

Block groups are the next level and comprise a cluster of census blocks. Block groups vary in size and generally contain between 600 and 3,000 people (average 1,500). Block group boundaries were originally defined to create population groups that are homogenous with regard to social and economic characteristics (US Department of Commerce et al., 1994). Residential segregation patterns in the US necessitate the use of such small area units for the study of small areas with high minority and/or immigrant populations.

Census tracts, the levels above block groups and composed of block groups, are small relatively permanent geographic areas within counties and comprise between 2,500 and 8,000 residents (average 4,000). Census tracts follow natural boundaries and are designed to be homogeneous with respect to population characteristics and living conditions (US Department of Commerce et al., 1994). Given the relative permanence of census tract boundaries, they are used

routinely by several Federal, state, and local agencies as administrative units for eligibility qualification and resource allocation (Subramanian et al., 2005).

Zip codes have commonly been used in health research. Zip codes differ markedly in definition and stability from census tracts and block groups. While census tracts and block groups are delineated to be homogenous units as described previously, zip codes are defined by the US Postal Service (USPS) for efficient mail delivery and can range in size from a single building to large areas that cross state boundaries. To overcome these area discrepancies, the 2000 US Census defined Zip Code Tabulation Areas (ZCTA) mapped to census blocks to replace zip codes. However, the current 5-digit ZCTA area may no longer correspond to the USPS 5-digit zip code area. The use of zip codes is further complicated by extensive modifications in the past 10 years, the creation of new zip codes, and the deletion of existing zip codes. In addition, census will not be providing any linkages between zip codes and ZCTAs. From a theoretical standpoint the use of zip codes as a level in health research is questionable (Krieger et al., 2002).

Counties—and the analogous parish (Louisiana), borough (e.g., Alaska), and county equivalent, hereafter referred to simply as county—are local levels of government, smaller than states but typically larger than cities or towns, with varying degrees of political and legal autonomy depending on the state. Counties provide a useful unit of analysis because they range from completely rural to metropolitan, providing a more representative geography of the US when compared to cities or metropolitan statistical areas. Counties incorporate a wider range of variability in levels of prosperity, demographics, and social and economic infrastructure. Additionally, educational, legal, and political institutions are generally shared within counties.

In contrast to counties, Metropolitan Statistical Areas (MSAs) consist of a large urbanized county or cluster of counties that have a high degree of social and economic

integration within that unit. MSAs are often used for programmatic purposes, including allocating Federal funds.

States—and the semi-analogous territories and tribes—are the primary political division in the US and have a large degree of autonomy from the Federal government. This autonomy can challenge Federal environmental laws and regulations, as the extent of implementation is highly variable across states, territories, and tribes.

### **Variations of Conventional Levels**

Given the appropriateness of a specific level and the data availability at that level, one can generate unique geographic and social contexts. For example, census block groups may be combined to create neighborhoods, or census tracts may be combined to create communities. State economic areas, economic sub-regions, and labor market areas can also be considered potential levels of analyses.

Other less commonly used census geographic entities can also be considered (US Department of Commerce et al., 1994). State economic areas (SEAs) are either a single county or a group of counties within a state, defined by economic similarities. An economic sub-region is a group of two or more economically similar counties that cross state lines. Both the SEA and economic sub-region should be reconsidered as important levels between the county and state. Labor market areas (LMAs) are one or more counties defined by commuting-to-work patterns and close economic ties and represent areas within which persons can reside and work within a reasonable distance and change jobs without changing their place of residence (US Department of Commerce et al., 1994).

## Appendix D: Inhalation Exposure to Outdoor Air Pollutants

Exposure assessment identifies who is exposed, as well as the level and pattern of exposure. Exposure assessment is based on data of the spatial and temporal patterns of air quality and population activity. The nature and complexity of an exposure assessment, including the spatial scale, is a function of the research question, characteristics of the exposure, multiple media sources of the pollutant, measurement methods, and policy.

Table D1 summarizes the elements of various types of exposure assessments. Except for proximity analysis, which provides only qualitative exposure comparisons for disparity analysis, all inhalation exposure assessments require estimates of the spatial pattern of air quality concentrations.

**Table D1: Types of Exposure Assessments**

Metric	Data Requirements	Analysis Approach	Exposure Accuracy	Resolution/ Extent
Residential proximity to emission sources	(a) Emission magnitudes/locations (b) Residential locations	GIS	Low (qualitative comparisons only)	High
Ambient concentrations at residential locations (monitoring)	(a) Air monitoring (b) Residential locations	GIS	Medium	Low
Ambient concentrations at residential locations (modeling)	(a) Emissions magnitudes/locations (b) Meteorology (c) Residential locations	Air dispersion modeling	Medium	Medium to High
Population exposure (monitoring)	(a) Personal monitoring (b) Residential locations (c) Demographics	Statistics	Low to medium	High
Population exposure (modeling)	(a) Air quality—monitored or modeled (b) Human activity (c) Residential locations	Population exposure modeling	Medium	Medium to High
Individual exposure	Personal monitoring	—	High	Low

## **Estimating Outdoor Air Pollutant Concentrations**

Outdoor air quality can be estimated in several different ways. The estimation procedure directly influences the level at which data are available, with monitoring data available at more aggregate levels such as counties and dispersion modeled data potentially available at the level of individual address.

Air monitoring provides actual measured concentrations at specific locations and times. One of the several limitations of monitoring, however, is that it is expensive. The spatial and temporal extent of the monitoring network is typically small, and/or resolution typically quite coarse. For example, California's South Coast Air Quality Management District recently completed the Multiple Air Toxics Exposure Study (MATES II). To represent air quality for the 15 million residents of Los Angeles, Orange, San Bernardino, and Riverside Counties, concentrations of 32 toxic air pollutants were measured over a year for 24-hour periods every sixth day at 10 fixed-site locations. Air toxics monitoring studies in other regions have measured concentrations at even fewer locations (e.g., eight in Detroit; six in Seattle, Tampa, and Providence; and five in Portland, OR).

Air dispersion modeling studies use data on emissions and meteorology to estimate air quality patterns according to physical principles of atmospheric physics and chemistry. These are generally more cost-effective than monitoring, and thus can be used to estimate air concentrations across larger areas and time periods with finer resolution, potentially to the latitude and longitude of an individual's address. For example, the air dispersion modeling portion of the MATES II study estimated concentrations of air toxics for every 24-hour period over a year at approximately 5,000 locations. US EPA's National Air Toxics Assessment

(NATA) national-scale assessment estimates annual average air toxics concentrations for each of the more than 60,000 US census tracts.

### **Estimating Exposure Concentrations**

Similarly, exposure concentrations may be measured directly with personal monitoring, but the cost generally precludes large study samples or monitoring periods longer than a few days. Exposure modeling combines air quality data (monitored or modeled) with data on population activity to estimate population exposure for various populations and demographic subgroups. Simple exposure modeling typically assumes populations are exposed to outdoor concentrations in the vicinity of their residences at all times, similar to the assumption of multilevel models where individuals in a given area share common exposures and experiences. However, more complex modeling uses population activity data from diary studies to account for people's movements among indoor and outdoor microenvironments and geographic locations. The NATA national-scale assessment uses such a model to estimate exposure concentrations for 10 age-gender groups in each US census tract.

### **Characteristics of the Pollutant**

Steeper concentration gradients require finer spatial resolution (e.g., census tracts or block groups) to accurately represent variations in exposure concentrations. Steeper gradients are expected for primary pollutants (i.e., those emitted directly from sources) than for secondary pollutants (i.e., those formed on the atmosphere from chemical transformation of precursor compounds, such as tropospheric ozone). Steeper gradients are also expected for large isolated emission sources (e.g., large power plants and large industrial facilities) than for sources that are

more dispersed (e.g., on-road motor vehicles). However, even for onroad motor vehicles there are steep concentration gradients within a couple hundred meters of a major roadway, so the contextual setting of the study population is also an important consideration.

## **Air Quality Research Goals**

### *Air Quality Standards*

The determination of whether an air quality standard has been violated requires the estimation of the maximum concentration in the modeling domain. When an air quality study is focused on a single, dominant emission source (e.g., a large isolated stationary emission source or a heavily trafficked, isolated roadway or intersection), concentration measurements may be made with a dense network of monitors spread over a very limited area. However, due to their limited spatial scope, such studies are not likely to be useful for health disparity analysis. For primary pollutants (i.e., steep gradients near emission sources), air dispersion modeling for maximum concentration determination is often done in a tiered manner, using coarse spatial resolution for the first tier to find the general area of highest concentrations and then using finer resolution on a subset of the domain in the second tier. The modeling receptors may be arranged in a gridded pattern or be associated with census subdivisions (e.g., internal points of tracts, block groups, and blocks).

### *Population Exposure Assessment*

The goal of a population exposure analysis may be to find the average exposure concentrations for various populations, stratified by geography and/or demography. In this case the choice of spatial resolution may depend on the resolution of the populations of interest, as well as the spatial variability of the air quality concentrations. The census tract has been

demonstrated generally to be an appropriate level to approximate the average exposure concentration for constituent ethnic groups adequately. Exceptions remain for highly segregated areas such as Harris, TX. In such cases census blocks or block groups may be more appropriate; however, some desired demographic details may not be available for the more spatially resolved data (see Table D2 and D3 below).

**Table D2: Percentage of Ethnic Group Members that Reside in US Census Blocks in which the Ethnic Group Fraction Is within 0.10 of the Ethnic Group Fraction of the Tract within which it Resides**

COUNTY	ETHNIC GROUP			
	White	Hispanic	Black	Asian
Los Angeles	74%	68%	67%	73%
Orange, CA	80%	67%	96%	80%
Brooklyn	81%	77%	85%	80%
Manhattan	88%	81%	86%	78%
Lorraine, OH (Cleveland)	88%	77%	83%	90%
Harris, TX (Houston)	60%	53%	58%	73%

Source: 1990 US Census data

**Table D3: Ethnic Composition of Six Selected US Counties**

COUNTY	TOTAL POP.	ETHNIC GROUP			
		White	Hispanic	Black	Asian
Los Angeles	8,863,164	41%	38%	11%	10%
Orange, CA	2,410,556	64%	23%	2%	10%
Brooklyn	2,300,664	40%	20%	35%	5%
Manhattan	1,487,536	49%	26%	18%	7%
Lorraine, OH (Cleveland)	1,412,140	72%	2%	25%	1%
Harris, TX (Houston)	2,818,199	54%	23%	19%	4%

1990 US Census data

Alternatively, the goal of a population exposure analysis may be to find the number of persons exposed above a given threshold. In this case, for steep concentration gradients, a finer

spatial resolution may be required so that spatial averaging does not mask exceedances of the threshold, e.g., census block rather than tract.

## **Multiple Media**

Multiple sources of air pollution data are an important consideration in determining the appropriate levels in the analytic design. Outdoor air pollutants are emitted by a variety of sources: major stationary sources, small industrial sources, commercial facilities, residences, on-road motor vehicles, and non-road mobile equipment. Residential proximity to the dominant source type for a given pollutant is often used as a surrogate measure for exposure to that pollutant when air quality concentration data are unavailable. Proximity to the source type can also be used as a surrogate for estimating spatial patterns of concentrations within the geographic subdivisions for which concentration data are available. For example, it is estimated that about 60% of CO exposure typically results from on-road vehicle emissions. If CO concentration estimates were available at the tract level, variations within the tract (e.g., at the block level) could be estimated based on traffic patterns within the tract.

## **Data Sources**

US EPA's NATA national-scale assessment: Modeling estimates of 1996 annual average census tract ambient concentrations and exposure concentrations of 33 air toxics. Modeling estimates of 1999 annual average census tract ambient concentrations and exposure concentrations are scheduled to be reported in Spring 2005.

South Coast Air Quality Management District (SCAQMD) MATESII study: Daily and annual average (1998-1999) air toxics monitoring (10 locations) and modeling estimates (2 km

spacing) for California's South Coast Air Quality Management District (Los Angeles, Orange, San Bernardino, and Riverside Counties).

Portland, OR Air Toxics Assessment (PATA): Daily and annual average (1999-2000) air toxics monitoring (five locations) and modeling estimates (block group and tract resolution).

SCAQMD Air Quality Management Plan: Episodic concentrations of tropospheric ozone (secondary), primary and secondary particles, and carbon monoxide (5 km spacing) for California's South Coast Air Quality Management District (Los Angeles, Orange, San Bernardino, and Riverside Counties).

Tropospheric ozone monitoring network: hourly concentration measurements at approximately 1,200 locations in the US.

Particulate matter (PM) network: 24-hour average concentration measurements every sixth day at approximately 1,000 locations in the US for PM<sub>10</sub> and approximately 1,000 locations for PM<sub>2.5</sub>.

### *Policy Level*

For criteria pollutants (ozone, carbon monoxide, nitrogen dioxide, sulfur dioxide, particulate matter, and lead) US EPA sets air quality standards. California has also set somewhat different air quality standards, generally stricter than US EPA's. States, and in some cases local air quality management districts, are responsible for attaining the standards within their jurisdictions by regulating emission sources. US EPA also promulgates emission regulations directly for some types of sources, such as onroad motor vehicles and some types of nonroad mobile equipment. US EPA also has promulgated control rules for some sources of non-criteria pollutants, i.e., air toxics.

## **Appendix E: Ambient and Drinking Water Quality**

### **Characteristics of the Exposure**

Exposure to environmental agents in water can occur through a variety of pathways, including: dermal contact, incidental ingestion, and inhalation during swimming in rivers, lakes, and oceans; ingestion of fish and shellfish from contaminated waters; direct ingestion of drinking water; indirect ingestion of water used to process foods; and inhalation of radon and other contaminants entering homes from groundwater or during showering.

The sources of the environmental agents also vary widely and include direct disposal of wastewaters, runoff from farms, groundwater contamination from land disposal, air emissions that subsequently deposit to surface waters, naturally occurring materials (e.g., geological arsenic and radiological deposits that leach into groundwater), and poorly constructed or maintained drinking water treatment and distribution systems.

One of the more studied water pathways—direct ingestion of drinking water—can require a variety of spatial and other levels of analysis, depending on the populations and contaminants being studied, which in turn are related to the physical and regulatory framework of drinking water systems. Briefly, US EPA regulates public water systems (PWS's), the US Food and Drug Administration (FDA) regulates bottled water, and state and local authorities—or no one—regulates very small or private supplies (i.e., private wells and systems that are too small to be considered PWS's). Each of these sources of drinking water can have widely differing—or no—standards and, thus, widely differing exposures and risks.

## Water Systems

US EPA's definition of a public water system (PWS) states that it is a system for the provision to the public of water for human consumption through pipes or other constructed conveyances, if such a system has at least 15 service connections or regularly serves at least 25 individuals at least 60 days out of the year. A PWS can be one of two basic types:

- Community water systems (CWS's) serve at least 15 service connections or 25 people year-round in their primary residences. Most residences, including homes, apartments, and condominiums in cities, small towns, and mobile home parks, are served by CWS's.
- Non-community water systems (NCWS's) are public water systems that serve the public but do not serve the same people year-round. There are two types of NCWS's:
  - *Non-transient non-community systems* (NTNCWS's) serve at least 25 of the same persons over 6 months per year (e.g., schools or factories that have their own water source).
  - *Transient non-community systems* (TNCWS's) serve at least 25 persons (but not the same 25) over 6 months per year (e.g., campgrounds or highway rest stops that have their own water source).

While the majority of people (about 65%) obtain most of their drinking water from PWS's, the rest are about evenly divided between drinking primarily bottled water and water from private wells (EPA, 2002; Lee et al., 2002). Furthermore, residents of rural or farm areas are more likely to drink private well water than municipal or bottled water. About half of

drinking water in the US is from surface water, and half is from ground water, although often several sources of water are blended for a given PWS.

The source of drinking water can have a significant effect on the type, concentration, and frequency of exposure to environmental agents. Given the variety of drinking water sources in some areas, therefore, exposure in these areas also can vary widely. In some cases, individuals in the same neighborhood or even the same household can experience different exposures if, for example, some individuals rely solely on municipal tap water (e.g., from a CWS) for their drinking water, some rely on bottled water, and some rely on water primarily from work (e.g., a NTNCWS). In a study of inter- and intra-ethnic variation in water intake among Tucson residents, for example, Hispanics reported differences such as much higher rates of bottled water consumption than did non-Hispanic whites (Williams et al., 2001).

Rural areas, where individual households often obtain their water from private wells, can experience very different exposures even when the wells are fairly close to each other because of local effects of landfills or other sources of contamination but also because the wells might be drilled to different depths and, thus, drawing from different aquifers. Exposure can also vary within a larger CWS. For example, contamination may occur close to the tap (e.g., lead in the home's faucet or pipes) rather than at the water source.

### **Monitoring Research Goals**

The estimation of risk or determination of whether a drinking water standard has been violated usually requires the sampling and analysis of water samples, often at the PWS level prior to distribution to homes (i.e., before the water enters the pipes to and in the homes) and sometimes at the water's source (e.g., river, aquifer). At times, sampling is conducted of

contaminated soils or other material (e.g., landfill contents) “upstream” of the water, and then modeling is conducted to estimate concentrations in the ingested water. Sometimes additional sampling or modeling of changes in concentrations after the water leaves the PWS or other system is required.

PWS’s and bottlers are required to conduct monitoring on a regular basis, though, as described above, the standards vary depending on the type of water supply. Also, sampling of water at the tap is not conducted regularly, and, thus, modeling often is needed.

In one example of modeling of trihalomethane (THM) concentrations from multiple sources of water in a PWS, a statistical model was constructed using sparse routinely collected THM measurements to obtain quarterly estimates of mean THM concentrations. The THM measurements were modeled using a Bayesian hierarchical mixture model, taking into account heterogeneity in THM concentrations between water originating from different source types, quarterly variation in THM concentrations due to factors such as precursor concentrations and the time the water has spent in the distribution system, and uncertainty in the true value of undetected and rounded measurements (Whitaker et al., 2004).

Using water monitoring or modeling data with demographic data at the census tract level can be difficult. For example, in a study of problems associated with collecting drinking water quality data for community studies, the task of evaluating water quality for each census tract was complicated by the fact that single census tracts were served by more than one system; each system usually had more than one well; and single wells had several episodes of testing for various contaminants (Whorton et al., 1988). This is a classic problem that the multilevel modeling approach described in “multiple contexts at the same level” can potentially address.

Other monitoring focuses more on the drinking water source. For example, McLaughlin et al. (2001) note how ambient water (not all of which is used as a source of drinking water) in predominately black counties received more than twice the mass of chemicals released to water per square kilometer than all counties.

Actual health outcome attributable to drinking water, as with other public health data, can be collected and aggregated at a variety of different levels, including individuals, census blocks, census tracts, counties, and health districts, depending on whether the data are obtained from national, state, or local disease surveillance registries or from specific surveys. Data used for Los Angeles County in 2000 and New York County in 2001 demonstrate the use of a variety of different geographic units—primarily health districts and neighborhoods, respectively—to demonstrate that Hispanics have higher rates of giardiasis and cryptosporidiosis, respectively, than other ethnic groups (Natural Resources Defense Council (NRDC), 2004).

Monitoring data are available from a variety of sources, including the American Water Works Association (AWWA), US EPA's Safe Drinking Water Information System (SDWIS), and state and local utilities reports. Such data usually are provided at the water supply level, which often equate to counties but also can equate to multiple counties or multiple units within counties. Thus, census tracts often combine to form congruent areas with water supply areas, yet a single census tract can also contain more than one water supply area or parts of areas.

## References

- Bhattacharya, J., DeLeire, T., Haider, S., Currie, J., 2003. Heat or eat? Cold-weather shocks and nutrition in poor American families. *Am. J. Public Health* 93 (7), 1149-1154.
- Bingenheimer, J. B., Raudenbush, S. W., 2004. Statistical and substantive inferences in public health: issues in the application of multilevel models. *Annual Review of Public Health* 25, 53-77.
- Braveman, P. A., 2003. Monitoring equity in health and healthcare: a conceptual framework. *J Health Popul Nutr* 21 (3), 181-192.
- Brown, M. J., Gardner, J., Sargent, J. D., Swartz, K., Hu, H., Timperi, R., 2001. The effectiveness of housing policies in reducing children's lead exposure. *American Journal of Public Health* 91 (4), 621-624.
- Bryk, A. S., Raudenbush, S. W., 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, Newbury Park, UK.
- Burnett, R., Ma, R., Jerrett, M., Goldberg, M. S., Cakmak, S., Pope, C. A., 3rd, et al., 2001. The spatial association between community air pollution and mortality: a new method of analyzing correlated geographic cohort data. *Environmental Health Perspectives* 109 Suppl 3, 375-380.
- Cassel, J., 1976. The contribution of the social environment to host resistance: the Fourth Wade Hampton Frost Lecture. *Am. J. Epidemiol.* 104 (2), 107-123.
- De Leeuw, J., Kreft, I., 2001. Software for Multilevel Analysis. In: *Multilevel Modelling of Health Statistics*. Leyland, A. H., Goldstein, H. (Eds.), Wiley, New York.
- Dietert, R. R., Etzel, R. A., Chen, D., Halonen, M., Holladay, S. D., Jarabek, A. M., et al., 2000. Workshop to identify critical windows of exposure for children's health: immune and

- respiratory systems work group summary. *Environ. Health Perspect.* 108 Suppl 3, 483-490.
- Diez-Roux, A. V., 1998. Research forum: bringing context back into epidemiology: variables and fallacies in multi-level analysis. *American Journal of Public Health* 88 (2), 216-222.
- Diez Roux, A. V., 2002. A glossary for multilevel analysis. *J Epidemiol Community Health* 56 (8), 588-594.
- Duncan, C., Jones, K., Moon, G., 1996. Health-related behaviour in context: a multilevel modelling approach. *Soc Sci Med* 42 (6), 817-830.
- Duncan, C., Jones, K., Moon, G., 1998. Context, composition and heterogeneity: using multilevel models in health research. *Soc Sci Med* 46 (1), 97-117.
- EPA, 2002. Drinking Water From Household Wells. EPA 816-K-02-003.
- Galobardes, B., Lynch, J. W., Davey Smith, G., 2004. Childhood socioeconomic circumstances and cause-specific mortality in adulthood: systematic review and interpretation. *Epidemiol. Rev.* 26, 7-21.
- Gee, G. C., Payne-Sturges, 2004. Environmental health disparities: a framework integrating psychosocial and environmental concepts. *Environ Health Perspect* 112 (17), 1645-1653.
- Goldstein, H., 1995. *Multilevel Statistical Models*. Arnold, London.
- Hox, J., Kreft, I., 1994. Multilevel analysis methods. *Sociological Methods and Research* (22), 283-299.
- Jones, K., 1991. Specifying and estimating multi-level models for geographic research. *Transactions of the Institute of British Geographers, New Series* (16), 148-160.
- Krieger, N., 1994. Epidemiology and the Web of Causation: Has Anyone Seen the Spider? *Social Science and Medicine* 39 (7), 887-903.

Krieger, N., Waterman, P., Chen, J. T., Soobader, M. J., Subramanian, S. V., Carson, R., 2002.

Zip code caveat: bias due to spatiotemporal mismatches between zip codes and US census-defined geographic areas--the Public Health Disparities Geocoding Project. *Am J Public Health* 92 (7), 1100-1102.

Krieger, N., Waterman, P., Lemieux, K., Zierler, S., Hogan, J. W., 2001. On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. *Am J Public Health* 91 (7), 1114-1116.

Langford, I. H., Bentham, G., McDonald, A. L., 1998. Multilevel modelling of geographically aggregated health data: a case study on malignant melanoma mortality and UV exposure in the European Community. *Statistics in Medicine* (17), 41-57.

Lee, S., Levy, D., Hightower, A., Imhoff, B., EIP FoodNet Working Group, 2002. Drinking Water Exposures and Perceptions among 1998-1999 FoodNet Survey Respondents. International Conference on Emerging Infectious Diseases, Atlanta, GA.

Luster, M. I., Dean, J. H., Germolec, D. R., 2003. Consensus workshop on methods to evaluate developmental immunotoxicity. *Environ. Health Perspect.* 111 (4), 579-583.

Macintyre, S., Ellaway, A., Cummins, S., 2002. Place effects on health: how can we conceptualise, operationalise and measure them? *Soc Sci Med* 55 (1), 125-139.

McEwen, B. S., 1998. Protective and Damaging Effects of Stress Mediators. *Seminars in Medicine of the Beth Israel Deaconess Medical Center* 338 (3), 171-179.

McLaughlin, D. K., Stokes, C. S., Nonoyama, A., 2001. Total and Race-specific Mortality in the U.S.: Inequality Across U.S. Counties. Department of Agricultural Economics and Rural Sociology and Population Research Institute, The Pennsylvania State University, University Park, PA.

- Morello-Frosch, R., Pastor, M., Jr., Porras, C., Sadd, J., 2002. Environmental justice and regional inequality in southern California: implications for future research. *Environ Health Perspect* 110 Suppl 2, 149-154.
- Natural Resources Defense Council (NRDC), 2004. Hidden Danger-Environmental Health Threats in the Latino Community.
- O'Neill, M. S., Jerrett, M., Kawachi, I., Levy, J. I., Cohen, A. J., Gouveia, N., et al., 2003. Health, wealth, and air pollution: advancing theory and methods. *Environ. Health Perspect.* 111 (16), 1861-1870.
- Rice, N., Jones, A., 1997. Multilevel models and health economics. *Health Economics* 6 (6), 561-575.
- Robinson, J. G., 2001. ESCAP II: Demographic Analysis Results. 1. Bureau of the Census, Washington, DC.
- Sampson, R. J., Raudenbush, S. W., Earls, F., 1997. Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy. *Science* 277 (August), 918-924.
- Schulz, A., Northridge, M. E., 2004. Social determinants of health: implications for environmental health promotion. *Health Educ Behav* 31 (4), 455-471.
- Singer, J. D., 1998. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics* 24 (4), 323-355.
- Subramanian, S. V., 2004. The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Soc Sci Med* 58 (10), 1961-1967.
- Subramanian, S. V., Chen, J. T., Rehkopf, D. H., Waterman, P. D., Krieger, N., 2005. Racial disparities in context: a multilevel analysis of neighborhood variations in poverty and

excess mortality among Black populations in Massachusetts. *Am J Public Health* 95 (2), 260-265.

Subramanian, S. V., Jones, K., Duncan, C., 2003. Multilevel methods for public health research. In: *Neighborhoods and Health*. Kawachi, I., Berkman, L. (Eds.), Oxford University Press, New York, NY, 65-111.

US Census Bureau, American Community Survey: What is the American Community Survey? Available: <http://www.census.gov/acs/www/SBasics/What/What1.htm> [accessed March 2 2005].

US Census Bureau, Standard hierarchy of census geographic entities. Available: <http://www.census.gov/geo/www/geodiagram.pdf> [accessed March 2 2005].

US Department of Commerce, Economics and Statistics Administration, US Census Bureau, 1994. Geographic Areas Reference Manual. Available: <http://www.census.gov/geo/www/garm.html> [accessed March, 2 2005].

Whitaker, H., Best, N., Nieuwenhuijsen, M. J., Wakefield, J., Fawell, J., Elliott, P., 2004. Modelling exposure to disinfection by-products in drinking water for an epidemiological study of adverse birth outcomes. *J. Expo. Anal. Environ. Epidemiol.*

Whorton, M. D., Morgan, R. W., Wong, O., Larson, S., Gordon, N., 1988. Problems associated with collecting drinking water quality data for community studies: a case example, Fresno County, California. *Am J Public Health* 78 (1), 47-51.

Williams, B. L., Florez, Y., Pettygrove, S., 2001. Inter- and intra-ethnic variation in water intake, contact, and source estimates among Tucson residents: Implications for exposure analysis. *J Expo Anal Environ Epidemiol* 11 (6), 510-521.

